A TEST FOR SPEECH DISCRIMINATION COMPOSED OF CNC MONOSYLLABIC WORDS (N.U. Auditory Test No. 4)

TECHNICAL DOCUMENTARY REPORT NO. SAM-TDR-62-135

January 1963

USAF School of Aerospace Medicine
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas

Project No 7755, Task No 59670

(Prepared under contract AF 41(657)-418
by Northwestern University, Evanston, Illinois)

FOREWORD

This report was prepared by the following personnel at Northwestern University

TOM W TILLMAN, Ph.D

RAYMOND CARHART, Ph D.

LAURA WILBER, M A

The authors are grateful to Richard Sweetman, who served as the talker in preparation of the magnetic tapes comprising N U Auditory Test No 4

ABSTRACT

The N. U. Auditory Test No. 4 is composed of two lists of 50 CNC monosyllabic words each that conform to the phonemic balance advocated by Lehiste and Peterson. The lists were given twice to three different groups of 16 subjects—those with normal hearing, those with conductive losses, and those with sensorineural losses. During each test, six presentation levels of ascending intensities were used, the total range being from −4 db to +40 db sensation level.

The three types of subjects evidenced articulation functions of the same shape, but the functions for sensorineurals were of gentler slope than for the other two groups. The discrimination scores for list I were slightly higher than for list II. During the retest, the discrimination scores improved slightly. Scores between lists as well as those from test to retest showed relatively high positive correlation. Therefore, the N. U. Auditory Test No. 4 seems to be a valuable tool for the measurement of phonemic discrimination.

This technical documentary report has been reviewed and is approved.

Robert B. Payne

ROBERT B. PAYNE
Colonel, USAF, MSC
Chief, Operations Division

A TEST FOR SPEECH DISCRIMINATION COMPOSED OF CNC MONOSYLLABIC WORDS
(N.U. Auditory Test No. 4)

I. HISTORICAL BACKGROUND

The technic of measuring discrimination for monosyllabic words has found many applications in audiology, both in research and in clinical practice. The pioneer efforts of Fletcher and his associates at the Bell Telephone Laboratories established the precedents and launched the concepts which persist today (9, 10), but it was the evolution of the PB-50 lists at the Psycho-Acoustic Laboratory in the early 1940's which was the direct impetus to development of present-day procedures in this phase of speech audiometry (5). The special features of the PB-50 lists are that they comprise 20 matched and nonduplicating compilations of 50 words each, that the 20 lists are relatively equivalent to one another, that a measure of discriminatory efficiency can be obtained with a single list, and that each sequence of 50 words approximates the phonetic balance of English.

These features appealed to clinicians in the military programs for aural rehabilitation which were active during World War II. Therefore, the PB-50 lists were quickly adopted in the military programs as a tool for assessing hearing impairment and for selection of hearing aids (2). The traditions for these clinical applications evolved in part from experience with large populations of hard-of-hearing servicemen (1). However, solidification of clinical practice took cognizance of the thinking of research groups, especially of those at the Psycho-Acoustic Laboratory and at the Central Institute for the Deaf. A noteworthy example of the benefits resulting from this early interaction between clinicians and theorists is the concept of the Index for Social Adequacy which was proposed in 1946 by Walsh and Silverman (20) and was elaborated by Davis in 1948 (4).

The intervening years have brought three major developments in this phase of testing for speech intelligibility.

First, measurement of discrimination for speech has become recognized as a basic diagnostic technic, and test materials have been developed in many languages. Clinicians have applied these materials to such diverse tasks as differentiating between conductive and sensorineural impairments (20), determining a patient's suitability for otologic surgery, diagnosing C.N.S. lesions in the auditory system (13), evaluating practical communicative impairment (3), and assigning medicolegal ratings of disability in hearing (19).

Second, measurement of discrimination for speech has become a procedure used commonly in research on the auditory functions of both normal hearers and persons with auditory impairments. Some of this research has sought to define the variables underlying the understanding of speech. Some of it has utilized measurement of speech intelligibility to obtain criterion data or background information on research subjects with whom other facets of hearing behavior were the prime topics under scrutiny. These two uses have brought us a substantial accumulation of knowledge regarding factors which affect precision in perceiving monosyllabic words and other types of speech material.

Third, both clinical and research experience with the original PB-50 word lists revealed limitations not originally appreciated. For example these lists include enough unfamiliar

Received for publication [illegible] Aug [illegible]

words to affect the performance of subjects with restricted vocabularies. Realization of such limitations triggered efforts to improve both the materials themselves and the methods by which they are presented. One outcome has been the development of several new tests. These tests fall into two categories. One type consists of new compilations of words which maintain the same pattern of phonetic balance used in the original PB-50 lists. The PB-K-50 test devised by Haskins (11) is one example of this approach, while the W-22 test recorded by Hirsh and his associates (12) is its best known example. The second type of test is one which employs new patterns of phonetic balance based on a revised criterion of what that balance should be. The two noteworthy products of this approach are the rhyme test by Fairbanks (8) and the Lehiste-Peterson CNC lists (14, 16). Both these sets of materials consist of monosyllabic word lists employing 50 items as the basic test unit. In this respect they preserve the pattern found in the original PB-50 lists and their successors.

At present, many research workers and clinicians are uncertain as to which of these various tests to choose for the purpose of speech audiometry. All these tests have three basic features in common: namely, they employ 50 words per block of test items, they include only monosyllables, and they incorporate some form of phonetic or phonemic balance. The question of choice must therefore rest on their merits as judged from theoretic considerations or as revealed by experience.

Some clinicians and researchers favor the W-22 test because so much clinical experience has been accumulated with it and because it has been subjected to substantial analysis of various types (6, 7, 15, 17). In fact, the W-22 test is used by the Veterans Administration and by the majority of audiologic clinics in the country. Other workers distrust the W-22 materials because these materials have proved relatively easy for most listeners and, therefore, do not differentiate sharply among minor deficits in ability in phonemic discrimination. Moreover, the W-22 test is available only on phonograph discs, and the commercial quality of these discs has at times been poor.

Some critics of the W-22 test have adopted other materials for purposes of measuring phonemic discrimination. Some have turned to the Fairbanks Rhyme Test. Champions of the rhyme test feel that it has a special advantage because it is very specific in the phonemic distinctions it requires the subject to make. The rhyme test has several limitations, however, which have kept it from being widely adopted. These disadvantages are that it only probes the discrimination of consonants, that its test items comprise a matrix that is too restricted, that the number of alternative versions available is not large (particularly since the test calls for multiple choice responses), and that the relationship of measures with the rhyme test and with PB-50 type material has been established with only moderate definitiveness.

Persons who doubt the wisdom of using the original PB-50 lists, the W-22 lists, or the Fairbanks Rhyme Test are attracted to the CNC monosyllabic word lists which Lehiste and Peterson developed. The latter lists have three major advantages. First, careful analysis of phonetic, phonemic, and linguistic considerations led Lehiste and Peterson to restrict their test items to monosyllabic words of the CNC variety—i.e., words composed of an initial consonant, a vowel nucleus, and a final consonant. Second, each list of 50 CNC words was edited to achieve the phonemic balance which characterizes the corpus of 1,263 monosyllabic words from which the CNC lists were drawn, rather than the phonemic structure of English as a whole. Thus, these lists have increased face validity in that they are representative of the specific type of material from which the test items themselves are derived. In the third place, Lehiste and Peterson took precautions to assure that the words they used were as familiar as was practical (14, 16). In view of these characteristics, the Lehiste-Peterson lists stand as materials which appear to have high potential as measures of speech intelligibility.

At the moment, the problem facing anyone wishing to employ the Lehiste-Peterson test is lack of information as to the reliability and interchangeability of the ten lists which comprise these materials. This limitation becomes particularly restrictive when their use is contemplated in a research project that requires many measurements of phonemic discrimination under closely allied circumstances.

The foregoing limitation was the impetus for development of N.U. Auditory Test No. 4. A situation developed where the research group in the Auditory Research Laboratory at Northwestern University needed two highly equivalent forms of a test of phonemic discrimination. The decision was made to use a test of the Lehiste-Peterson variety. Two lists could have been chosen arbitrarily for this purpose from among the ten published by Lehiste and Peterson. Before these two lists could have been used insightfully, however, it would have been necessary to subject them to extensive experimental study for the purpose of defining their reliability and interchangeability. Since such a substantial task was unavoidable, it was practical to develop two new lists and then subject these to the aforementioned analysis. The latter course was chosen because Lehiste and Peterson had found it necessary in composing their ten lists to deviate slightly from their original plan for phonemic balance. It seemed better to start an investigation of reliability and interchangeability with two lists which conformed precisely to this pattern than to work with less perfect examples of it. The decision was therefore made to prepare two new lists and to designate them as N.U. Auditory Test No. 4.

The discussion which follows reports on the development of these two new CNC lists, describes the preparation (on magnetic tape) of six alternate forms of each list, and summarizes experimentally derived facts as to reliability and other features of these new materials.

2. NATURE OF N.U. AUDITORY TEST NO. 4

The CNC word lists

As already mentioned, the two parent lists developed in evolving N.U. Auditory Test No. 4 were made to conform rigorously to the phonemic balance advocated by Lehiste and Peterson. They obtained this pattern of phonemic balance by extracting all the 1,263 monosyllables in the consonant-vowel-consonant class which are listed by Thorndike and Lorge (18) as occurring at least once per million words. Lehiste and Peterson determined the frequency of occurrence of each initial, medial, and final phoneme in these 1,263 monosyllables. They specified that each phoneme should occur in a single list of 50 consonant-vowel-consonant words with the same incidence as it exhibited in the total list of 1,263 words. They then proceeded to construct ten lists of CNC[1] words according to this plan. These lists were published in 1959 (14). The 500 words comprising the ten lists were selected from the original pool of 1,263 CNC monosyllables.

The first step in developing CNC lists conforming to the Lehiste-Peterson pattern was to make the tabulation which appears in table I. This table shows the number of times each phoneme is to be used in order to preserve, as closely as one can in a 50-word list, the distribution which characterizes the 1,263 CNC words that Lehiste and Peterson abstracted from the Thorndike and Lorge compilation.

The second step was to evolve two mutually exclusive groups of 50 words that contained exactly the distribution of phonemes appearing in table I. These two new groupings of words are reported in table II. They have been designated as list I and list II of N.U. Auditory Test No. 4. Moreover, all but five of the 100 words, thus assembled, were chosen from the 500-word pool comprising the original ten Lehiste-Peterson lists. The remaining five words all appear in the larger pool of 1,263 words.

The third step was to randomize each of the new parent lists six times. The randomizations which were thus at hand were subsequently recorded on magnetic tape.

[1] The designation CNC rather than CVC was chosen by these authors to stress the fact that the phonetic elements identified as vowels are actually syllabic nuclei in monosyllables of the type selected.

TABLE I

The proportions (p) of incidences of phonemes which constitute the Lehiste-Peterson pattern of phonemic balance for CNC words and the number (N) of inclusions of each phoneme required to develop a single list of N U Auditory Test No 4

Initial consonant			Vowel nucleus			Final consonant		
Sound	p	N	Sound	p	N	Sound	p	N
p	0642	3	i	0832	4	p	0564	3
b	0658	3	ɪ	1116	5	b	0264	1
t	0578	3	eI	0942	5	t	1102	6
d	0594	3	ɛ	0744	4	d	0778	4
k	0658	3	æ	1038	5	k	0818	4
g	0420	2	ɑ	0864	4	g	0392	2
m	0584	3	ʌ	0592	3	m	0542	3
n	0460	2	ɔ	0626	2	n	1054	5
f	0452	2	oU	0736	4	ŋ	0208	1
v	0182	1	U	0222	1	f	0310	2
θ	0118	1	u	0586	3	v	0288	1
ð	0080	0	aI	0278	1	θ	0240	1
s	0680	3	aU	0736	4	ð	0090	0
z	0032	0	ɔI	0126	1	s	0564	3
ʃ	0354	2	ɜ	0562	3	z	0190	2
r	0736	4				ʃ	0200	1
l	0736	4				ʒ	0018	0
tʃ	0316	2				r	0628	3
dʒ	0332	2				l	1032	5
h	0606	3				tʃ	0318	2
w	0476	2				dʒ	0200	1
hw	0156	1						
j	0150	1						
	1 0000	50		1 0000	50		1 0000	50

$N = \frac{100p}{2}$ rounded to the nearest integer

Before describing the procedure employed in recording N U Auditory Test No 4, a further comment must be made regarding the composition of the two lists When they were being composed, no special effort was made to equate them to one another or to the Lehiste-Peterson lists in terms of word familiarity This factor was not originally considered by Lehiste and Peterson (14) Since they have subsequently revised their lists, however, to eliminate some of the more unfamiliar items in their original compilation, it is interesting to note how N U Auditory Test No 4 conforms in this regard to the revision Table III allows this comparison by reporting the proportions of words in each test which fall into each of seven classes of familiarity It is immediately apparent from the table that the relative distribution of test words among these classes is about the same for the two N U lists as for the average of the Lehiste-Peterson revision It is furthermore apparent that these several lists not only include a sizeable fraction of very common words, but also encompass a wide range of familiarity

Recording the CNC lists

The apparatus and procedures employed in making the parent recording of each list in

TABLE II

Alphabetical arrangement of CNC monosyllabic words comprising lists I and II of N.U. Auditory Test No. 4

List I		List II	
bean*	met	bite	merge*
boat	mode*	book*	mill
burn	moor	bought	nice*
chalk	nag*	calm	numb
choice	page	chair*	pad*
death*	pool	chief	pike
dime*	puff*	dab*	pur*
door	rag	dead*	rain
fall*	raid	deep	read*
fat*	raise*	faith	room
gap	reach	fir	rot
goose*	sell	gaze	said
hash*	shout*	gin	shack*
home	size	goal	shawl
hurl*	sub	hate*	sick
jail	sure	haze	soap
jar	take	hush	thought
keen	third	juice	tool
king	tip	keep	towel
kite	tough	keg	turn
knock	vine	learn	voice
laud	week	live	wag
limb	which	loaf	white
lot	whip	lore	witch
love*	yes	match	young

*Also in original PB-50 lists

the N.U. Auditory Test No. 4 are described as follows.

A condenser microphone (Western Electric 640AA, Serial No. 1059) was mounted in free air within an acoustic booth (Industrial Acoustics Company, Inc., Model 1202) with inside dimensions of 5.3 by 6.0 by 6.5 feet. The microphone was electrically coupled to a cathode follower circuit (Western Electro-Acoustical Laboratory Preamplifier, Type E), which in turn fed a condenser microphone complement (Western Electro-Acoustic Laboratory, Type 100D E). The output of the condenser microphone complement was led through a speech audiometer (Grason-Stadler, Model 162), whose output fed a single channel of a two-track magnetic tape recorder (Ampex, Model 351-2).

A male talker, 23 years of age, who uses General American dialect, was chosen to speak the test materials. He was experienced in the monitored live-voice technic of speech audiometry. In addition, he was given adequate practice with the N.U. Auditory Test No. 4 before the final recording of its two lists was undertaken. During the latter recording he

TABLE III

Distributions according to frequency of usage of the CNC monosyllables in N.U. Auditory Test No. 4 and in the revised Lehiste-Peterson test

Familiarity rating	N.U. Auditory Test No. 4		Average per list in Lehiste-Peterson revision
	List I	List II	
Among most common 500 words	11	13	10.9
Among next most common 500 words	11	6	8.3
More than 100 occurrences per million words	2	1	0.8
50 through 99 occurrences per million words	3	7	7.7
25 through 49 occurrences per million words	10	5	8.3
10 through 24 occurrences per million words	9	7	7.8
1 through 9 occurrences per million words	4	11	6.2

spoke into the 640AA microphone at an intensity just sufficient to cause a deflection to −3 db on the VU meter of the speech audiometer. This deflection was achieved while a standard carrier phrase, "Say the word ___," was being said. The test item followed the carrier phrase naturally as part of a continuous utterance without an attempt to monitor the test item to any particular level. The talker presented items at intervals of five seconds. An assistant helped the talker to maintain the proper timing by tapping him gently on the back every five seconds. The randomized sequence of each list designated as form A was employed. Both lists were recorded on the same day, on the same tape recorder, but on separate magnetic tapes. In addition, a 1000 cps sinusoid was recorded ahead of each list as a calibration tone. The level of this tone was adjusted so that it, too, caused a deflection to −3 db on the VU meter of the Grason-Stadler unit. In order to avoid saturating the tape, the recording level was adjusted to produce a deflection of −20 db on the VU meter of the tape recorder.

Recording spondaic words

In order to have available material by the same talker which could be used to establish speech reception thresholds, it was decided to prepare two randomizations of 30 spondaic words. The two series were recorded with the same equipment and at the same level re a 1000 cps calibration tone as that used in preparing the magnetic tape carrying the N U Auditory Test No. 4. The time interval between successive spondees was five seconds. In this instance, however, the carrier phrase was omitted and each spondee was monitored individually to the standardized meter reading of −3 db.

Preparation of alternate forms of lists

The recorded version of N U Auditory Test No. 4 was assembled in five additional forms, or randomizations, of each list. These were designated as Forms B, C, D, E, and F.

The procedure employed was very simple. Each master list was copied six times. Since the original recording had used the word order for form A, one copy was retained without change. The other five copies were reorganized into the remaining five orders of word scrambling. The first step in preparing another form was to separate its component items by cutting the magnetic tape into short segments, being very careful to leave 4.5 seconds of "silence" as the leader to each item. The list was then reassembled in the new order by the laborious process of splicing its 50 items in the requisite succession. As a final step, the companion forms of the two lists plus a section of the 1000 cps calibration tone were spliced together. The resultant combination of two lists and the calibration tone constituted a finished magnetic tape. For example, the magnetic tape carrying form B consisted of the calibration tone, list I-B and list II-B.

The six magnetic tapes thus assembled were the specific materials employed in the analysis of N U Auditory Test No. 4, which is described in the remainder of this report.

3 METHOD OF EVALUATION OF N U AUDITORY TEST NO. 4

Administration of lists at selected presentation levels

The equivalence between list I and list II, their test-retest reliabilities, and other characteristics of N U Auditory Test No. 4 were evaluated on three groups of subjects. These subjects were 16 persons with normal hearing, 16 with conductive impairments, and 16 with sensorineural losses.

The basic procedure in gathering data was to see each subject twice. During each of these sessions, both lists were administered to the subject six times. Each successive presentation was at a higher intensity level, but both lists were administered at one level before the transition to the next higher one. The purpose in this stairstep procedure was to allow plotting of the articulation function for each list with as little contamination from practice and learning as possible. This goal was achieved by having the first presentation at such a low level that only the most audible items were understandable The fact that these words had been received correctly on the first presentation had no effect on the next score since these words would have been understood at the higher level, irrespective of the earlier successful contact with them Of course, a new group of words became understandable at this higher level It was the additional success with these latter words which was responsible for improvement in the articulation score after the transition from the first level to the second one The same principle, now true for a second group of new words, operated at the third presentation level, and so on

This procedure did not eliminate guessing when a subject perceived part of the phonemes in a word, but it kept him from being alerted to the existence of specific words as members of the test vocabulary.

The sections which follow describe these subjects and procedures more fully, and then review the findings which emerged from analysis of the articulation scores thus obtained

Subjects

Normal hearing subjects were procured from the student population at Northwestern University, whereas the two groups of hearing-impaired subjects were drawn from persons who had been seen in the Northwestern University Hearing Clinics

Thirteen men and 3 women comprised the normal hearing group Their mean age was 22 9 years and they ranged in age from 18 to 31 years The prime requirement for selection was that the subject have one ear which was better than 10 db hearing level as determined by a pure tone screen over the 125 to 8000 cps range.* The nontest ear was not held to this standard since all measurements were to be made monaurally

Only subjects known to suffer from clinical otosclerosis were included in the conductive loss group Preliminary selection in terms of degree of hearing loss was made from clinical records, but suitability was confirmed at the time the subject came in for his first experimental session At this point his pure tone thresholds were established for both air and bone conduction His speech reception threshold (SRT) was also measured Only subjects whose SRT's were at a hearing level between 20 and 58 db (re· 22 db SPL) were retained Moreover, a prospective subject was rejected if his bone conduction level in the 250 to 4000 cps range showed a maximum loss of 25 db at two frequencies or more than a 25 db loss at one frequency The patient's unoperated ear was selected when such a choice was necessary. Finally, the subject was required to be at least 18 but not more than 50 years old at the time he participated in the study As it turned out, the mean age for the group was 38 2 years Seven subjects were male and 9 were female

The 16 subjects selected for the sensorineural loss group were persons who had experienced progressive hearing loss during adulthood They were all drawn from the diagnostic categories of sensorineural loss of unknown origin, familial sensorineural loss, and early presbycusis These types were selected in an effort to avoid subjects with special problems in discrimination such as are encountered in Meniere's disease and VIIIth nerve tumor As with the conductive loss group, final decision to include a subject with sensorineural loss was based on his audiometric performance at the time of his appointment for the first experimental session The requirements were that his air and bone conduction audiograms interweave, that his better ear serve as the test ear, and that this ear exhibit a speech reception threshold at a hearing level between 20 and 58 db re 22 db SPL. Each subject was required to be in the age range of 25 through 50 years The mean age of this group was 42 9 years, and the sex distribution was 12 women and 4 men

Test procedures

Each subject participated in two test sessions As already mentioned, these sessions were identical except for the fact that pure tone audiometry was omitted in the second session and that the presentation level for one pair of the CNC lists was altered as described below

The sequence during the initial session began for normals with the pure tone screening test The sequence began for both hearing loss groups with the establishment of pure tone thresholds for air conduction and bone conduction by the Hughson-Westlake technic A Maico MA-2 audiometer calibrated to conform to the NBS norms was used as the test instrument in all these measures

The monaural speech reception threshold was next measured in each ear The subject was familiarized with the list of spondees in advance of any testing with them This familiarization was accomplished by having him repeat the spondees as these were read to him by the tester in a face-to-face situation Actual determination of the SRT involved presenting the first of the recorded spondees at a sensation level of approximately 10 db and then descending in 2 db steps presenting four words at each step until the lowest level was reached at which either two out of the first three or two out of the four spondees were understood This level was designated as the speech reception threshold The materials used were the two recorded lists of spondees mentioned earlier This material was reproduced through a speech audiometer (Grason-Stadler, Model 162) calibrated to conform to the ASA norm wherein 22 db re 0 0002 microbar represents "0" db hearing level A tape recorder (Ampex, Model 601) delivered the test material to the external input of one of the channels of the speech audiometer In every instance, the level of the 1000 cps calibration tone which accompanied the test material was set so that the VU meter of the speech audiometer registered 0 db

The next step was to obtain discrimination scores with the 12 scramblings of N U Auditory Test No 4 The speech reception threshold for the subject's test ear constituted the base of reference for setting all sensation levels at which these CNC materials were administered The orders in which the two parent lists were presented were counterbalanced, half the subjects starting with list I and the other half with list II Each list was presented at six sensation levels namely, 0, 8, 16, 24, 32, and 40 db Both lists were presented at one level before the next higher level was entered Within this framework of alternating the two lists and cumulating the presentation level, the various forms of each list were employed in random order As a precaution against obtaining response through the nontest ear, 60 db of effective masking was delivered to this ear in those few instances in which the presentation level called for in the test ear exceeded the SRT in the nontest ear by 40 db or more As already explained, the plan of progressing from a low sensation level (0 db SL) to a high one (+40 db SL) was chosen to minimize contamination of discrimination scores by correct guesses based on knowledge of the test items The rationale is that an item which is heard correctly at one level will also be understood at higher levels, so that a subject's score should not be changed at one level by the fact that he had already identified some of the items at lower levels The initial session terminated when the twelfth list had been administered at the +40 db sensation level

After an interval of one to two weeks, the retest session was undertaken Here the first step was to establish monaural SRT's anew The same ear was retained as the test ear, but presentation levels for CNC materials were established in terms of the retest SRT for this ear Again, the two lists were presented at one level before the next level was employed, and again, the pattern of progressive increase was used This time, however, the first presentation was at −4 db sensation level, and the sequence thereafter was 0, +8, +16, +24, and +32 db sensation level. Moreover, each subject received the lists and forms in exactly the same order as he had during the initial session except that the forms originally presented at +40 db sensation level were now given at −4 db sensation level[2] Finally, the

[2]The −4 db sensation level was substituted on the retest because as reported later subjects were found to do so well at the +40 db level that it was not yielding data of great value

same rule was followed for using masking to avoid hearing in the nontest ear. The retest terminated when both lists had been presented at +32 db sensation level.

Articulation functions for normal hearing subjects

Table IV summarizes the data obtained with normal hearing subjects during the initial test session, while table V reports the companion information for the retest session. These tables report medians, means, and average deviations of discrimination scores separately for each list at each presentation level. The four articulation functions that can be derived from these two tables are illustrated in figure 1, where the functions are plotted from the means of scores obtained for each list during each experimental session.

Figure 1 reveals that the two lists yielded essentially the same slope of articulation function during the initial session and the retest run. These curves are displaced from one another slightly on the horizontal scale, but their parallelism is unusually good. They fit the same configurational pattern with definitiveness, thus giving one confidence that they are valid descriptions of the way performance

TABLE IV

Discrimination scores obtained with N.U. Auditory Test No. 4 for subjects with normal hearing during the first test session (scores represent percent of items correctly repeated)

Sensation level of presentation*	List I			List II		
	Median	Mean	Average deviation	Median	Mean	Average deviation
0	27	28.3	8.0	21	23.0	9.9
+ 8	78	75.8	7.5	73	69.4	11.5
+16	94	93.3	3.4	91	92.9	3.0
+24	100	99.0	1.2	100	98.6	1.5
+32	100	99.9	0.2	100	99.8	0.5
+40	100	99.9	0.2	100	99.8	0.5

*Mean SRT = 25.4 db SPL

TABLE V

Discrimination scores obtained with N. U. Auditory Test No. 4 for subjects with normal hearing during retest session (scores represent percent of items correctly repeated)

Sensation level of presentation*	List I			List II		
	Median	Mean	Average deviation	Median	Mean	Average deviation
− 4	7	7.0	4.8	2	3.6	3.9
0	30	34.4	11.2	20	23.1	8.7
+ 8	77	78.1	8.6	79	75.8	9.0
+16	95	94.6	4.1	94	94.8	4.1
+24	100	99.3	1.1	100	99.5	0.8
+32	100	99.5	0.8	100	99.5	0.9

*Mean SRT = 24.6 db SPL

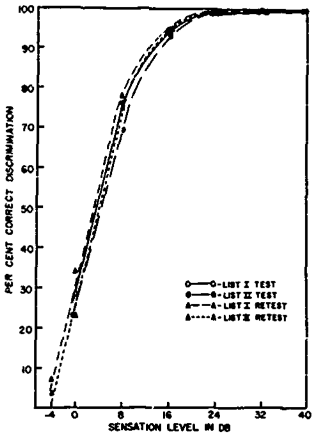

FIGURE 1

Mean articulation functions yielded by normal-hearing group for lists I and II during both test and retest sessions

by normals on N U Auditory Test No 4 is related to presentation level

The outstanding feature of this pattern is that it appears to represent a linear function which undergoes saturation at higher signal intensities The lower segment of the curve is linear and has a slope of about 6% increase in discrimination score per decibel increase in presentation level This linear segment terminates at sensation levels of about 9 db, where discrimination scores approach 80% The upper portion of the function is a curvilinear progression wherein scores improve less and less per decibel of stimulus elevation until an asymptote, characterized by almost perfect discrimination, is reached This asymptotic response is achieved when the presentation has reached a sensation level of +24 db

The foregoing features exhibit themselves in another guise when one considers the variability of scores at different presentation levels This variability was found to be large within the linear portions of the function. Variability decreased progressively and dramatically once the level was high enough to saturate the performance with correct responses The situation is exemplified by the average deviations reported in tables I and II At sensation levels of 0 db and +8 db, which lie below the threshold of saturation, the average deviations range from 7 5 to 11 5% At the +16 db sensation level, which is at the point of partial saturation, the average deviation drops to 4 1 or less As the asymptote is approached—i e, at +24 db sensation level—the average deviations become so small that they are less than 1% At this point, variability among scores for normal hearing subjects is undoubtedly assignable to occasional errors due to lack of attention and other secondary factors

Although, as previously stressed, the general shape of the articulation function did not change from list I to list II nor from test to retest, minor differences did appear in consequence of both variables Before discussing these effects in detail, however, it is desirable to note how, if at all, the basic pattern of the articulation function for N U Auditory Test No 4 was modified when the test was administered to the subjects with conductive loss and to the subjects with sensorineural loss

Articulation functions for subjects with conductive hearing loss

Tables VI and VII report the medians, means, and average deviations of the discrimination scores yielded by the 16 subjects with conductive loss employed in the present study Figure 2 plots the four articulation functions derived from these means and thus summarizes performance on each list during each experimental session

The most noteworthy observation to be made regarding these data is that they duplicate very closely the picture yielded by normal hearing subjects Each of the four functions

TABLE VI

Discrimination scores obtained with N. U. Auditory Test No. 4 for subjects with conductive hearing losses during first session (scores represent percent of items correctly repeated)

Sensation level of presentation*	List I			List II		
	Median	Mean	Average deviation	Median	Mean	Average deviation
0	22	22.9	9.4	23	20.1	9.4
+ 8	74	71.2	12.2	65	65.4	12.2
+16	94	92.6	5.5	95	92.0	5.5
+24	100	98.0	1.2	100	99.0	1.2
+32	100	99.6	0.6	100	99.6	0.6
+40	100	99.7	0.2	100	99.8	0.2

*Mean SRT = 61.3 db SPL

TABLE VII

Discrimination scores obtained with N. U. Auditory Test No. 4 for subjects with conductive hearing losses during retest session (scores represent percent of items correctly repeated)

Sensation level of presentation*	List I			List II		
	Median	Mean	Average deviation	Median	Mean	Average deviation
− 4	8	9.7	6.7	2	7.1	7.5
0	36	30.5	13.4	22	25.3	13.7
+ 8	75	74.1	9.2	74	70.3	10.9
+16	96	93.9	4.9	96	94.0	4.8
+24	100	98.0	2.5	100	98.5	2.0
+32	100	99.0	1.2	100	99.4	1.0

*Mean SRT 61.3 db SPL

exhibits the configuration of a linear function which reaches saturation. Moreover, the results at each sensation level are numerically very close for the two groups. True, the conductives show slightly greater discrepancy between functions from test to retest and from list I to list II. Also, as gaged by the average deviations, the homogeneity of responses at a given sensation level was slightly poorer for the subjects with conductive loss, but, again, average deviations are restricted sharply as the saturation region is invaded. Again, too, the slope of the linear segment of each function is nearly 6 % per decibel change in presentation level.

Articulation functions for subjects with sensorineural loss

Tables VIII and IX, plus figure 3, summarize the results for subjects with sensorineural loss. One notes immediately that these data have two close parallels with the data for the other two groups of subjects and two major discrepancies therefrom.

As was true for the other two groups, the four sets of data for sensorineural subjects agree with one another closely. True, they deviate slightly in the absolute values involved, but they yield articulation functions which

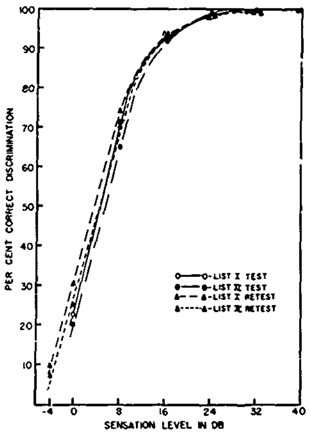

FIGURE 2

Mean articulation functions yielded by conductive-loss group for lists I and II during both test and retest sessions

are highly equivalent to one another in shape and location on the chart. Thus, any one of the four configurations is characteristic of the relationship between increase in presentation level and improvement in phonemic discrimination which this group exhibits

The second feature of similarity is that, again, the pattern bespeaks a saturation curve In this instance, the linear segment continues up to the point where the mean discrimination score is about 65% the contour being nonlinear thereafter

The first discrepancy from findings with normal hearers and cases with conductive loss is that the articulation functions for sensorineural cases are much more gradual in slope and the transition to nonlinearity occurs at a lower mean discrimination score In consequence of these two features, the nonlinear upper segments of these functions do not flatten off to full saturation within the range of presentation levels employed in the present study Mean discrimination scores at the +40 sensation level are approximately 92%, and it appears likely from the contours of the four functions that the saturation asymptote would be reached at a mean discrimination score of 95% or slightly less (fig 3)

TABLE VIII

Discrimination scores obtained with N U Auditory Test No 4 for subjects with sensorineurol hearing losses during first session (scores represent percent of items correctly repeated)

Sensation level of presentation*	List I			List II		
	Median	Mean	Average deviation	Median	Mean	Average deviation
0	11	12.5	6.8	7	7.9	5.6
+ 8	31	35.4	13.8	27	30.5	15.4
+16	60	62.0	13.2	52	55.8	14.1
+24	81	79.0	12.5	76	75.5	10.8
+32	90	87.6	8.5	89	86.6	9.6
+40	93	92.0	6.2	96	93.0	6.9

*Mean SRT = 5[illegible] 4 db SPL

TABLE IX

Discrimination scores obtained with N U Auditory Test No 4 for subjects with sensorineural hearing losses during retest session (scores represent percent of items correctly repeated)

Sensation level of presentation	List I			List II		
	Median	Mean	Average deviation	Median	Mean	Average deviation
− 4	3	4.1	3.7	2	2.5	2.3
0	11	14.6	8.5	5	8.1	6.8
+ 8	39	42.6	18.7	26	35.8	17.7
+16	64	66.4	14.2	59	59.9	12.6
+24	85	81.5	12.1	82	78.9	12.3
+32	91	88.8	8.0	92	89.0	8.0

*Mean SRT 7.8 db SL

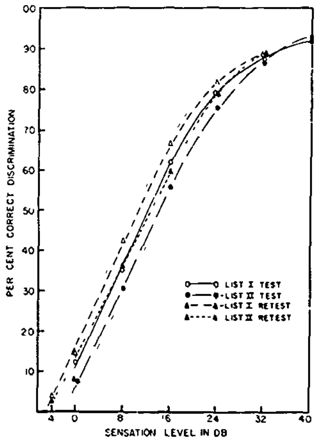

FIGURE 3

Mean articulation functions yielded by sensorineural group for lists I and II during both test and retest sessions

The second unique feature of the sensorineural group is that its discrimination scores were much more variable at a given presentation level than were those of the other two groups This fact is revealed by the average deviations which were appreciably greater for the sensorineural subjects at most presentation levels Obviously, these results indicate that the sensorineural subjects were a less homogeneous group in their behavior than were either the normal hearers or the subjects with conductive loss seen in this study This fact is at least partially responsible for the more gradual slope to the articulation function of the sensorineural group

Comparison of articulation functions for the three groups

The parallelism and equivalence in functions for normals and conductives is so great that it appears fully justifiable to accept a combination of the data for these two groups as the basis for a generalized description of the articulation function characterizing N U Auditory Test No 4 This description is given graphically as curve A in figure 4, which averages all scores for these 32 subjects as tested with both lists during the two experimental sessions This composite function

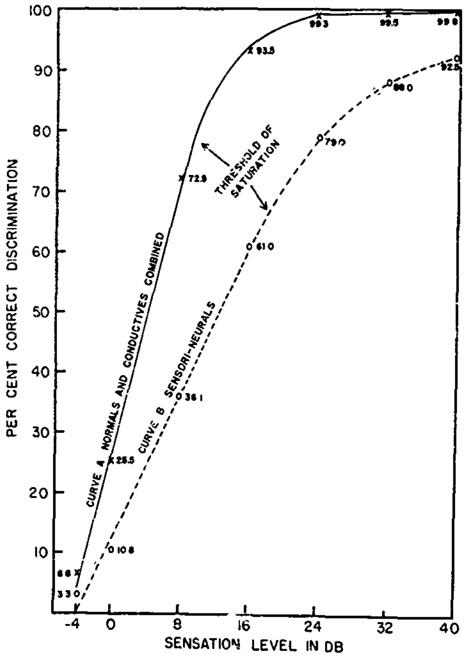

FIGURE 4

Mean articulation functions for N U Auditory Test No 4 derived by averaging test and retest data for both lists

stands as the best estimate from present data of the relation to be expected when using N U. Auditory Test No 4 if the subject's capacity for phonemic discrimination is not defective, albeit he may exhibit hypoacousis The salient features of this function are that the slope of its linear segment is 5 6% per decibel, that no evidence of saturation appears until the discrimination score approaches 75%, and that saturation is virtually complete at the +24 db sensation level.

It is against this frame of reference that one should evaluate the articulation function for N U Auditory Test No 4 as yielded by the 16 subjects with sensorineural loss This comparison requires averaging this group's scores for both lists and for the two test sessions, as has been done in order to derive curve B of figure 4

The relations reviewed earlier stand out sharply when one compares the two curves in figure 4 These two curves clearly belong to the same family, but the curve for sensorineurals is characterized by a linear segment with more gradual slope (3% per decibel as opposed to 5 6% per decibel) Moreover, the transition to nonlinearity occurs with a lower mean discrimination score for the sensorineurals (about 65%) than for the other two types of subjects (about 75%) These two features combine to yield a curve for sensorineurals which does not achieve full saturation—i e, reach its asymptote, even at a presentation level of +40 db sensation level This latter level is 16 db greater than the +24 db sensation level at which the curve for the other two groups exhibits full saturation. A final difference involves the discrimination scores associated with the plateau, or asymptote of saturation Ruling out occasional chance errors due to factors such as lapses of attention normal hearers and subjects with conductive loss were capable of achieving perfect discrimination scores so that the asymptote for their mean function is essentially at 100% Although the mean asymptote for the sensorineural group was not reached in the present study, extrapolation of curve B in figure 4 suggests that it would not have been found to exceed 94 or 95%. This observation is not an unexpected finding in view of the demonstrated clinical fact that some sensorineurals do not achieve perfect phonemic discrimination at any level of word presentation Actually, it is rather astonishing that the mean discrimination at high levels for this group is as good as it was found to be Clearly, the group exemplified sensorineural hypoacousis uncontaminated by dysacousic factors which would have independently disturbed phonemic discrimination

The relationships illustrated in figure 4, and just discussed, have particular importance insofar as they bear on the purposes for which N.U Auditory Test No. 4 was developed: namely, to achieve a test of phonemic discrimination which would be useful in determining the slope of a person's articulation function. They reveal that N.U Auditory Test No. 4 satisfies the two requirements which such a test must possess First, it must incorporate an articulation function whose slope is linear over a substantial segment of its course, so that the value of the slope may be measured with reasonable precision Second, the degree of the slope must vary from one type of subject to another since there would be no reason for undertaking to measure the slope of the articulation function if it were invariant from person to person

Before one proceeds to use N U Auditory Test No 4, however, he must know whether its two parent lists are acceptably equivalent to one another and whether unfortunate practice effects accrue as the lists are repeated The assessment of these two factors, as revealed by the data already described, constitutes the next two sections of this report

Equivalence of list I and list II

Figures 1, 2, and 3 reveal that the articulation functions for list II are displaced slightly to the right of the functions for list I These displacements suggest that the two lists differ systematically in such a manner that list II requires slightly greater intensity than list I for subjects to achieve the same discrimination score, particularly throughout the linear segment of the function Since this relationship is found in the results obtained with all three types of subjects, it becomes necessary to evaluate its magnitude and significance

Table X summarizes the pertinent facts for each group of subjects separately and for the combination of all three groups The table records the algebraic means of differences between discrimination scores for the two lists at each sensation level, as well as the probability associated with the statistic yielded when the sign test was applied to the data for that sensation level The instances in which these probabilities were equal to or less than 05 for a two-tailed test are marked with an asterisk to help the reader evaluate the significance of the findings

TABLE X

Differences between means of discrimination scores for lists I and II at the several presentation levels, and the probability associated with each difference as derived from the sign test (test and retest data combined)

Sensation level of presentation	Normal hearers		Conductive loss group		Sensorineural loss group		All groups combined	
	$L_I - L_{II}$	P	$L_I - L_{II}$	P	$L_I - L_{II}$	P	$L_I - L_{II}$	P
−4 (retest only)	3.4	.09	2.6*	.02	1.6	.10	2.5*	.0005
0	8.3*	.0001	4.0	.06	5.7*	.001	6.0*	.0001
8	4.2	.06	5.0*	.04	5.7*	.03	5.0	.0003
16	0.1	.10	0.3	.10	6.4*	.03	2.3	.10
24	0.1	.10	−0.8	.10	3.1*	.05	0.8	.10
32	−0.2	.10	−0.2	.10	0.4	.10	0.0	.10
40 (test only)	0.2	.10	0.0	.10	−1.0	.10	−0.3	.10

*P < .05

The relationships which appear in table X are clean cut. The largest differences between means occurred at sensation levels of 0 and +8 db for normal hearing and conductive loss cases. The sensorineural group is characterized by large differences not only at these two sensation levels but also at +16 and +24 db. The foregoing relationships present an integrated picture when one remembers that normals and conductives ceased to show linear increase of discrimination scores with increase in presentation level above +8 db sensation level and that sensorineurals did not cease to show this linear increase until the +24 db sensation level. Thus, the sensation levels at which large differences appeared for each group are the sensation levels where the linear relation between discrimination score and intensity of presentation was maintained. As soon as the presentation became high enough to start evoking saturation responses, the mean difference between scores with the two lists decreased to a very small value. The relationship may be highlighted in another way. There are ten instances reported in table X in which the mean difference between scores for a single group is greater than 2.5%. All ten are associated with the linear segment of an articulation function and in all cases the performance was better on list I. The sign test yielded a probability which was equal to or less than .05 in seven of these instances. None of the differences associated with levels evoking saturation responses exceeded 8% and none of these differences was statistically significant at confidence levels more stringent than 10%. These facts support the conclusion that a real difference between the two lists exists but that it is apparent only when the level of presentation is low enough to keep the response on the linear segment of the subject's articulation function.

Thus, other things being equivalent, the score for list I may be expected to be better than for list II. The difference to be expected, however, is not large. As can be seen in table X, the average difference for all groups combined is only 6.0% at 0 db sensation level and 5.0% at +8 db sensation level. Statistical evaluation of these differences by means of the sign test yields statistics whose probabilities are less than .0003. These results give one high confidence in the reality of the difference between lists, but they also highlight the fact that the difference is not numerically great when evaluated in terms of the slope that characterizes the linear portion of the articulation functions. This slope, it will be recalled, is 5.6% per decibel for normals and conductives and 3.0% for sensorineurals. Therefore, the systematic difference between lists in discrimination score is equivalent to a displacement of between one and two decibels on the intensity scale Expressed thus, the difference is no greater than the margin of uncertainty which is associated with establishing a speech reception threshold, which is the base of reference from which sensation level is computed

These findings, plus the observation that the two lists have articulation functions with equivalent linear slopes, lead to the conviction that the two lists are interchangeable within the margins of clinical precision These findings also warrant the conclusion that, provided one precaution is observed, it is legitimate to use the two lists in research designed to establish the slopes of articulation functions This precaution is that only the several scramblings of a single list may be used to define the slope of a particular articulation function In other words, the alternate list may be substituted only when a new set of conditions is introduced for which a new function is to be explored

Another way of assessing the comparability of the two lists is to consider the correlations between the discrimination scores obtained with them To this end, coefficients of correlation were determined by the product moment method for those sensation levels lying on the linear segment of the articulation function Use of this method seems appropriate, since one may presume that the scores encompassed by the linear segment of the articulation function lie on an equal interval scale Table XI reports these correlations not only for each group of subjects separately, but also for all subjects together at those sensation levels where combining their results was justifiable

The general conclusion which one may reach from examining table XI is that lists I and II are sufficiently interdependent to warrant their being considered as alternate forms of a single exploratory tool All correlations are positive and most of them approach or exceed 80

On the whole, the relationship revealed by the coefficient of correlation is least strong for the group of normal hearers, and at 0 db sensation level during the initial test session *r* is only 44 This particular correlation may be attributed to the fact that the normal listeners used as subjects were unfamiliar with speech audiometry at the outset of the study It will be recalled that 0 db sensation level was the very first condition used during the initial experimental session These subjects apparently experienced some uncertainty of response until they became acclimated to the test procedure In consequence, the correlation during the very first test condition was probably poorer than it would have been if they had been familiar with the basic technics of speech audiometry Confirmation of this interpretation is found in two facts First, all other coefficients for the normal group are higher than 44 by a noteworthy degree Second, the

TABLE XI

Coefficients of correlation (Pearson r) between lists I and II at sensation levels where articulation functions were judged to be linear

Sensation level	Normal hearers		Conductive loss group		Sensorineural loss group		All groups combined	
	1st session	2d session	1st session	2d session	1st session	2d session	1st session	2d session
−4	—	71	—	90	—	78	—	84
0	44	78	82	83	71	78	77	85
8	72	79	81	76	93	89	93	92
16	—	—	—	—	87	90	—	—
24	—	—	—	—	90	92	—	—

two sets of data for the subjects with hearing loss yielded high correlation at 0 db sensation level even during the first session The latter subjects had all taken several discrimination tests during earlier clinical examinations and were familiar before the experiment started with the basic technics of speech audiometry

It still must be reiterated, however, that normals also tended to exhibit slightly lower correlations between lists under the other test conditions than did the other two groups At −4 db sensation level, for example, r = 71 for normals and 90 and 78, respectively, for conductives and sensorineurals This trend is probably evidence that skill in phonemic discrimination itself was more homogeneous in the normal hearers under study, so that, in their case, chance factors exerted a greater relative influence on the statistical outcome The importance of this observation is that when the transition to the testing of subjects with hearing loss is made, the relationship to be expected between the two lists is enhanced This situation increases the confidence with which these materials may be used as alternate forms when carrying on discrimination testing

In conclusion and summary, it is pertinent to point out that for all three groups combined, the coefficients of correlation were 84 at −4 db, 77 and 85 at 0 db, and 92 and 93 at +8 db sensation level These results allow one to consider lists I and II of N U Auditory Test No 4 as highly equivalent when used to explore phonemic discrimination along the linear segment of the articulation function

Test-retest relationships

Comparison of the results for the first session with those obtained during the second session reveals two relationships of importance First, discrimination scores improved slightly during the second session as gaged by the means of scores obtained at sensation levels where the articulation function was linear Second, test-retest reliability was good

In considering the first of these relations, it is not possible to say whether the prime reason for the improvement which occurred in mean discrimination obtained during the second session was practice per se or was familiarization with items in the test lists Table XII presents the facts, however, which allow one to evaluate the magnitude and statistical significance of the change at the five sensation levels employed during both sessions Mean discrimination scores for normal hearers and for the subjects with conductive loss improved during the second session at sensation levels of 0 and +8 db by amounts which ranged from 3 2 to 6 3% By contrast, improvements were less than 2% at the +16 db and less than 1% at the +24 and +32 db sensation levels Results for the sensorineural group were somewhat different The largest shifts, ranging from 2 9 to 6 3%, occurred at sensation levels of +8, +16, and +24 db Improvements at 0 and +32 db sensation level were less than 2%

Only in four of all the foregoing instances did application of the sign test yield statistics whose probabilities were equal to or less than 05, and these four results were distributed haphazardly over three sensation levels and the three types of subjects The positiveness of the aforementioned trend is greatly enhanced, however, when one considers the combined data for all three groups Here, the sign test yielded statistics which were significant at the 0002, 002, and 005 levels of confidence at sensation levels of −4, 0, and +8 db, respectively Of course, it must also be noted that the mean improvement on the retest was less than 5% in each of these three instances The mean shifts in score for the +24 and +32 db sensation level were very small and the probabilities associated with these differences were greater than 10 in each case

One may reason from these results that the effects of practice and familiarity did not constitute major contaminants to discrimination scores at least within the range of the exposures employed in the present study The mean changes observed were less than would have been caused by a one decibel increase in presentation level Thus, the magnitude of the observed shift is not sufficient to destroy one's confidence in the discrimination score obtained when a particular scrambling of either

TABLE XII

Differences between means of discrimination scores for the test condition and those for the retest condition at the several presentation levels, and the probability associated with each difference as derived from the sign test (data for lists I and II combined)

Sensation level of presentation	Normal hearers		Conductive loss group		Sensorineural loss group		All groups combined	
	T_I-T_{II}	P	T_I-T_{II}	P	T_I-T_{II}	P	T_I-T_{II}	P
0	−3 2*	04	−6 3*	03	−1 7	09	−3 7*	0002
8	−4 3	10	−3 7	10	−6 3*	02	−4 8*	002
16	−1 6*	02	−1 3	10	−4 2	10	−2 4*	005
24	0 6	10	0 2	10	−2 9	10	−1 1	10
32	0 3	IND	0 4	10	−1 8	10	−0 4	10

*P = 05

IND Insufficient number of differences to allow sign test

TABLE XIII

Coefficients of correlation (Pearson r) between test and retest at sensation levels where articulation functions were judged to be linear

Sensation level	Normal hearers		Conductive loss group		Sensorineural loss group		All groups combined	
	List I	List II	List I	List II	List I	List II	List I	List II
0	62	40	57	66	81	54	73	65
8	60	79	74	78	80	86	88	91
16	—	—	—	—	76	76	—	—
24	—	—	—	—	81	79	—	—

list must be used a second time It also seems proper to conclude that a sequence of repetitions of N U Auditory Test No 4 at progressively higher presentation levels may safely be used in future experimentation without fear of inducing major shifts in the discrimination score because exposures are cumulative, provided the amount of exposure does not exceed that employed in the present study This conclusion, of course, presumes that the several randomizations of the lists will be used, since it is not safe to assume that the relations would remain as here described if the word order were not being altered from one presentation to the next

The question as to the reliability of lists I and II is best approached through examination of the coefficients of correlation between test scores and the retest scores Table XIII reports these coefficients for each group of subjects, for each list separately, and at the several sensation levels below the region of saturation

The picture which emerges is gratifying All coefficients for separate groups of subjects are positive. The lowest coefficient to emerge is 40 for list II administered to normals at the 0 db sensation level The other coefficients for the several groups of subjects ranged upward to 86 for the sensorineurals at the +8 db sensation level on the retest Moreover, when the data for all three groups were combined, the coefficients of correlation between test and retest were found to be 65 and .91 at

0 db and +8 db sensation levels, respectively. The various values reported in table XIII are of the order of magnitude which is generally considered to indicate good test-retest reliability. One may therefore conclude that in this respect, too, N.U. Auditory Test No. 4 has revealed satisfactory stability.

Other relationships of interest

A variety of other relationships may be observed in the data at hand. Two of these are worthy of brief discussion.

The first of these is the effect exerted by subject sophistication. As pointed out earlier, phonemic discrimination improves slightly with exposure to test materials, and list I yields slightly higher discrimination scores than list II. One would have expected these two effects to have accumulated particularly unfavorably when list II was the first test given to audiologically naive subjects. Conversely, the two effects should have tended to cancel one another when the opposite order of presentation was employed. Such an outcome occurred with the normal hearing subjects. These people, it will be recalled, lacked prior experience with speech audiometry. The eight normals who received list II first at 0 db sensation level in the initial session averaged 12.0% poorer on this list than on list I. The remaining eight normals averaged 1.5% better on list II, which they heard after having been familiarized with the testing technic through use of list I. Evidence of similar interaction between familiarization and test list is apparent to a lesser degree at other levels and with the other groups of subjects. The practical implication of this observation for the clinician is that the minor differences in the difficulty of the two lists can be largely counteracted by using list I in the initial stages of an exploration and by reserving list II for the final stages.

A second observation is that individual differences in subject response appeared. This trend is revealed by the fact that discrimination scores for each group of subjects tended to align themselves similarly at different sensation levels encompassed in the linear segment of the articulation function. This fact is clearly marshaled in the coefficients of correlation which emerged when adjacent presentation levels for the several lists and the several groups of subjects are considered. These coefficients are reported in table XIV. One notes that the coefficients are all positive and that they range from .50 to .91. Values of this magnitude indicate that systematic individual differences are incorporated in the data under study.

TABLE XIV

Coefficients of correlation (Pearson r) between adjacent presentation levels but restricted to consideration of sensation levels where articulation functions were judged to be linear

	Normal hearers		Conductive loss group		Sensorineural loss group			All groups combined	
	−4 vs 0	0 vs +8	−4 vs 0	0 vs +8	−4 vs 0	0 vs +8	+8 vs +16	−4 vs 0	0 vs +8
First session									
List I	—	.50	—	.65	—	.70	.82	—	.78
List II	—	.65	—	.77	—	.70	.91	—	.78
Second session									
List I	.63	.61	.73	.65	.73	.50	.85	.68	.69
List II	.75	.40	.80	.65	.68	.86	.86	.73	.71

One factor which undoubtedly contributes to these systematic individual differences is that minor errors of measurement could not help but be involved in determining the SRT's for the various subjects. These SRT's, in turn, became the bases from which the sensation levels for presenting the N.U. materials were specified. To illustrate, two persons with identical articulation functions would tend to reveal systematic displacements between their discrimination scores if their tests were not administered from identical sensation levels because their SRT's had not been specified with complete accuracy.

This factor, while undoubtedly involved in the results reported in table XIV, is not the only one having an influence. The coefficients of correlation for test-retest comparisons offer clear evidence that true variation in ability to make phonemic distinctions also appeared within each of the three groups of subjects (table XIII). The fact that these correlations are all positive and are relatively high does more than give one confidence in the reliability of N.U. Auditory Test No. 4. Results of this kind could have occurred only if each group tested also included a sufficient range of abilities to allow parallelism between scores obtained in the two test sessions to appear. Clearly, the correlations reported in table XIII would not be expected to have resulted purely from inaccuracies in estimating the SRT's for the two sessions, since errors of measurement for these SRT's would not be expected to be aligned in the same order in the two instances.

4. CONCLUSION

The foregoing discussions lead one to believe that N.U. Auditory Test No. 4 is a valuable addition to the array of materials available for measurement of phonemic discrimination. Persons with normal hearing, with conductive losses, and with sensorineural losses (considered both as separate groups and in the composite) yielded results with this test which allow one to state that list I and list II are closely equivalent forms. Both lists have good reliability as gaged by correlations obtained when test and retest data were compared. Differences between the two lists and familiarity with the test materials exerted only second-order influences on discrimination scores. Systematic differences among individual subjects appeared in all three groups under study. These differences are most clearly apparent when lists are administered at sensation levels where discrimination scores are linearly related to signal intensity. Hence, N.U. Auditory Test No. 4 appears to be a particularly promising tool for determining the slope of the linear portion of an individual's articulation function.

REFERENCES

1. Carhart, R. Speech audiometry in clinical evaluation. Acta Otolaryng. (Stockh.) 41:18-42 (1952).

2. Carhart, R. Tests for selection of hearing aids. Laryngoscope 56:780-794 (1946).

3. Carhart, R. The determination of hearing loss. Information Bulletin (Audiology) IB 10-115. Department of Medicine and Surgery, Veterans Administration. Washington: U.S. Government Printing Office, 1960.

4. Davis, H. The articulation area and the social adequacy index for hearing. Laryngoscope 58:761-778 (1948).

5. Egan, J. P. Articulation testing methods II. OSRD Report 3802 (1944).

6. Elpern, B. S. Differences in difficulty among the CID W-22 auditory tests. Laryngoscope 70:1560-1565 (1960).

7. Elpern, B. S. The relative stability of half-list and full-list discrimination tests. Laryngoscope 71:30-35 (1961).

8. Fairbanks, G. Test of phonemic discrimination: The Rhyme Test. J. Acoust. Soc. Amer. 30:596-601 (1958).

9. Fletcher, H. Speech and hearing. New York: Van Nostrand, 1929.

10. Fletcher, H. Speech and hearing in communication. New York: Van Nostrand, 1953.

11. Haskins, H. A phonetically balanced test of speech discrimination for children. Master's Thesis, Northwestern University, 1949.

12. Hirsh, I. J., H. Davis, S. R. Silverman, E. G. Reynolds, E. Eldert, and R. Benson. Development of materials for speech audiometry. J. Speech Hearing Dis. 17:321-337 (1952).

13. Jerger, J. Audiological manifestations of lesions in the auditory nervous system. Laryngoscope 70:417-425 (1960).

14. Lehiste, I., and G. E. Peterson. Linguistic considerations in the study of speech intelligibility. J. Acoust. Soc. Amer. 31:280-286 (1959).

15. Oyer, H. J., and M. Doudna. Structural analysis of word responses made by hard of hearing subjects on a discrimination test. AMA Arch. Otolaryng. 70:357-364 (1959).

16. Peterson, G. E., and I. Lehiste. Revised CNC lists for auditory tests. J. Speech Hearing Dis. 27:62-70 (1962).

17. Silverman, S. R., and I. J. Hirsh. Problems related to the use of speech in clinical audiometry. Ann. Otol. 64:1234-1245 (1955).

18. Thorndike, E. L., and I. Lorge. The teacher's word book of 30,000 words. New York: Columbia University Teachers College, 1952.

19. Veterans Administration. Rating of disability from hearing impairments. Extension 8-B, Schedule for rating disabilities, 1945 ed. Washington: Veterans Administration, 1956.

20. Walsh, T. E., and S. R. Silverman. Diagnosis and evaluation in fenestration. Laryngoscope 56:536-555 (1946).